

**B.Sc. (I. T.) Sem. - V (CBCS - 2015 Course) : WINTER - 2018**

**SUBJECT: DATA WAREHOUSING AND DATA MINING**

Day: Monday  
Date: 03/12/2018

**W-2018-1083**

Time: 10.00 a.m. to 01.00 p.m.  
Max. Marks: 60

**N.B.:**

- 1) Attempt **ANY SIX** questions.
- 2) Figures to the right indicate **FULL** marks.
- 3) Draw appropriate figures **WHEREVER** necessary.

- Q.1** a) Explain Data Warehouse in detail. (06)  
b) Explain Role of Business Intelligence in Data Warehouse. (04)
- Q.2** a) Define the term 'Data Mining'. Also explain four classes of task involved in a Data Mining process. (06)  
b) List and explain the features of Any Four data warehousing tools, popularly used in the industry. (04)
- Q.3** Suppose that a data warehouse for Big University consists of the following four dimensions: *student*, *semester* and *instructor* and two measures *count* and *avg grade*. When at the lowest conceptual level (e.g., for a given student, course, semester and instructor combination), the *avg grade* measure stores the actual course grade of the student. At higher conceptual levels, *avg grade* stores the average grade for the given combination. (10)  
i) Draw a snowflake *schema* diagram for the data warehouse.  
ii) Starting with the base cuboid [student; course; semester; instructor], what specific OLAP operations (e.g., roll-up from semester to year) should one perform in order to list the average grade of CS courses for each Big University student.
- Q.4** What is ETL? Explain the concept of ETL in detail with complete processing. (10)
- Q.5** a) Differentiate between normal view and materialized view. (05)  
b) What do you mean by Aggregate functions? Explain with suitable examples. (05)
- Q.6** A database has five transactions. Let  $\text{min\_sup}=60\%$  and  $\text{min\_conf}=80\%$  (10)

TID	Items-bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Find all frequent itemsets using Apriori algorithm and generated association rules along with their values for support and confidence.

- Q.7** Briefly compare the following concepts. You may use an example to explain your points. (10)  
i) Snowflake schema, fact constellation, starlet query model  
ii) Data cleaning, data transformation, refresh  
iii) Enterprise warehouse, data mart, virtual warehouse
- Q.8** Write short notes on: (10)  
a) Data model  
b) Data Mart